IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Applicant: E. Amitay

Serial No.:   10/743,158

Attorney Docket No.: IL920030045US1

Filed: December 22, 2003

Title: SEARCH ENGINE FOR ENHANCED WEB
                      INDEX

Examiner: Pham, M.

Group Art Unit: 2167

Declaration of Inventors

I, Einat Amitay, being first duly sworn, deposes and says:

1.   I hereby declare that I believe that I am the original, first and sole inventor of the subject matter which is claimed herein for which a utility patent is sought on the invention described and claimed in the above-identified application, and I have personal knowledge of all the facts herein stated.

2.   At all relevant times, I worked on this matter while in the employ of the assignee IBM and all right, title and interest in and to the invention and this application and any issuing patent are owned by IBM. This Declaration is, therefore, made on behalf of IBM and at its request.

3.   Pursuant to 37 C.F.R. 1.132, I hereby present this Declaration in order to present to the Examining Attorney additional pertinent factual information to assist him in distinguishing the herein claimed invention from the prior art.

4.   Whenever you do a "search," especially on the Internet, you must differentiate between "all relevant documents" and "all documents that contain the query." The set of all documents that have the query includes many documents that are not relevant, and the set of all relevant documents includes many documents that do not contain the query. The overlap area contains documents that are both relevant and also contain the query. To optimize searching you want to optimize this overlap area.

5.   What I do to optimize this overlap area is to "change" (or enhance) the document being searched by adding query words. This expands the possibility of a query finding relevant documents. In the index, the copy of the document is expanded (or enhanced) to also include the query words. In this manner, there is an increased possibility that the document will be found when someone in the future makes a query with the same query word as previously used by someone searching for the same (or similar) item.

6.   Prokoph (US Patent Publication 2002/0091671), on the other hand, is taking an entirely different approach. He is concerned that document size is getting too big and cannot be processed in an efficient manner. So, he reduces the size of the documents being searched thereby to reduce the total amount of data. His goal is to try and filter out data that is likely to be irrelevant. This is accomplished by replacing the document text by using abstracts. A problem with this approach, however, is that you are reducing the portion of the document being searched, so you necessarily increase the possibility of missing relevant documents.

7.   My goal is to increase the likelihood of finding relevant documents. Prokoph's goal is to filter out possibly irrelevant information so as to reduce the total size of the documents being searched so as to hopefully reduce the number of documents being returned in a search to possibly achieve a more manageable number of hits for the search engine and the user.

8.   Cole (US Patent 6,571,239) is creating a controlled vocabulary, whereas I use free language. He starts with a keyword index where he optimizes the language for keywords. If a user enters a query, and it appears on the keyword index, then the query keyword points to a document. My method does not involve optimizing language of the queries. Instead I add the query string directly into the document. In this way, if a future user presents this same query, the document comes up quickly among the list of relevant documents. Cole is instead interested in perfecting the language of the query and does nothing to change the documents so as to increase the likely of a particular document being retrieved in a search.

9.   As an initial point of reference the principal claim 37 herein reads.

        A method comprising:

                receiving user queries;

                searching an enhanced web index of documents with user queries, and

                wherein said enhanced web index containing document information and text,

                metadata and anchor text; and

adding information from at least some of said user queries to said enhanced web index.

10.   New independent claim 65 herein reads

An improved method for searching document indexes containing at least complete document information, wherein the improvement comprises:

adding information from at least some of user queries to said complete document index to create an enhanced index of at least complete documents and some information from said user queries; and

searching said enhanced index of complete documents with user queries upon receipt of an user inquiry.

11.   As explained above, the Invention deals with searching in complete documents on the Internet via queries. To increase the possibility of finding relevant documents, I embed information from past queries into the document so that, if someone uses the same inquiry in the future, there is a greater possibility that the document will be found. The cited prior art does not add anything to complete documents that are being searched. In one case, searching is done of an index of abstracts of the documents. Another reference maintains an enhanced list of keywords. Neither, however, discloses or suggests adding information from the query to the document so as to increase the ability of finding it in future searches.

12.   My invention is that the "index" of complete (not partial) documents is supplemented with "query words" created by users. Thus, these query words become part of the index of complete documents and will more readily identify the relevant document when future queries are made with this language. Claims 65, 36 and 56 specify clearly the type of index that is covered by the herein Invention.

13.   As explained in the herein instant application, "As many people have discovered, finding things on "The Web" can be easy, but only if the user knows the right terms to use to do the search. The right terms are those used by the designers of the web pages. This makes finding non-specific items difficult." (paragraphs 7 *et seq*) In other words, when a search is done of a web index, too much information is found and much of it is really not relevant. This is because the search engine can necessarily only search terms in the web index.

14.   In order to resolve this problem, I discovered that the search can be more meaningful and accurate if the web index is enhanced to include the text of user inquiries. "... there is a

significant amount of information in user's queries about how users view the items for which they are searching. In accordance with a preferred embodiment of the present invention, the query words may be joined to the information in the index, thereby increasing the ways in which an item may be described." (Paragraph 16 of the subject application). What this means is that the words of the query are inserted into the index of complete documents (e.g. a web index) so that they are there for future searches. These words are the words people are typically using to locate information. By placing them in the document index itself, when future users use the same term in their query, they come up with the more relevant portion of the index more quickly because the keywords of the query are already in the index. Hence it is the step of - adding information from at least some of said user queries to said document index – coupled with then using the thus enhanced index for searching - which is new and novel and distinguishes it from the prior art.

15.   With this background, Prokoph may be considered. He does not search an enhanced index of documents nor does he add information from the user queries to the enhanced index. Herein lies the difference between the claimed invention and Prokoph and it would not have been obvious to a person skilled in the art.

16.   The distinction between us is very simple. Propkoph has a method involving "... retrieving a document to be indexed, generating a document extract from the document, wherein the document extract comprises a portion of the document, and decomposing the document extract into tokens. The tokens are then stored in a search index, wherein a search engine accesses the search index to retrieve information satifying a search query." (Paragraph 22). What I do is inherently distinct. In my method, as disclosed in the published application, the index enhancer 16 may add terms to index 18 based on users' queries submitted to search engine 14.

17.   Prokoph is creating document extracts which are then searched, whereas I insert the actual query of the user into the actual index of complete documents.

18.   I am searching an Internet based index of complete documents, whereas Prokoph is searching a specially created index of only abstracts of documents. He does not search complete documents.

19.   Prokoph, by his own admission, creates a search index which consists of key words that have been created by parsing the documents previously found on the Internet. This is

what he searches, whereas I search an enhanced index containing complete document information. My index has full data, whereas Prokoph has only keywords created from parsed documents. These indexes are very different.

20.   Therefore, the two key differences between Prokoph and me is that he not only does not search an index of complete documents but he says to not search them and also he does not add information from user queries to the index of complete documents being searched.

21.   The Examiner also relied on Cole. This patent does not deal with indexes of complete documents. He does mention updating with user queries, but that is again something that was known.

22.   Reference is made to Cole's Abstract, wherein he defines his invention. "This invention provides methods, apparatus, system, and article of manufacture which solve the problem of mismatch between the keywords employed by a user in making a query and those assigned by the manual or automatic classification system stored in the system's keyword index." Their system deals with keyword indexes and not with an index of complete documents.

23.   Cole recognizes that "Existing keyword search engines for information repositories typically have two components. The first component may be described as a system for classifying a corpus of documents or other objects, such as images. The result of this process is a set of indices or similar data structures that associate keywords or terms with the documents or other objects. The second component provides a means for a user of the search engine to express a query. This component analyzes the query and uses the data structures provided by the first component to provide a set of objects which are deemed to be relevant to the user's query." (Column 1, lines 12 – 23). Their invention centers on this second component. Instead of being concerned about the index that contains the data, they are concerned about the index that contains the keywords for queries.

24.   What Cole is doing is modifying "the associations between objects in the database and keywords in the index, based on keywords supplied by the user during a search session." (Column 2, lines 5 -10). There is no modification of the database of complete documents. Instead they modify the keyword index.

25.   It must be appreciated that there is a significant difference between a keyword index and the actual database being searched. The keyword index is a repository of keywords used

during queries which have proven to be useful. They are kept for use in future queries. The actual database is left unchanged by Cole. It has no modification.

26. Cole's invention is summarized quite clearly in Column 2, lines 45 -67. "In an embodiment of the present invention, the system records the initial keyword(s) input by the user and holds them until the user is either satisfied or gives up. If in a query session the user is satisfied with the object(s) retrieved from the repository, the system associates the initial keyword(s) with the retrieved object(s). This facilitates the object's retrieval by the same user or subsequent users who input the same keywords. The association of new keywords with data objects is implemented in different ways, depending on whether modification of the master keyword index is allowed or desirable. Two alternative example embodiments are described below. Alternative A details a case in which the keyword index is modified directly. This is feasible, for example, when a single service or application controls the interaction between user and repository, end to end. Alternative B details the case in which the master keyword index is not modified. This is the case when the process interacting with the user does not have permission to change the master index. This occurs when, for example, only experienced librarians may have authorization to modify it. In this case, new keywords are stored in an auxiliary index. An external process merges both the master and auxiliary indices before returning the results to the user."

27. Cole is clear that only the keyword index is modified. "…Alternative A details a case in which the keyword index is modified directly. … Alternative B … is the case when the process interacting with the user does not have permission to change the master index. … In this case, new keywords are stored in an auxiliary index. An external process merges both the master and auxiliary indices before returning the results to the user." I do not know how it could be stated more clearly that they are adding information from queries to the keyword index and not to the index of complete documents.

28. Cole teaches "FIG. 4 describes an example showing an overall flow of the system. The user submits a query (401) which is matched against the Keyword Index (405) and against the Auxiliary Index (410) under Alternative B. If this is the user's first query in a session (412), the query (and its statistics) is stored in the Updating module (413). The matched keywords are used by the system to retrieve objects associated with them (425). The objects (or their description) are then displayed to the user (430). If the user enters a response

APPLICANT(S):   E. Amitay
SERIAL NO.:   10/743,158
FILED: December 22, 2003
Page 7

which indicates satisfaction (440), the first query stored (in step 413) is parsed into keywords (445). Each keyword is associated with an object, and optionally, statistics of date and usage are updated for each association (450). Under Alternative A, the Master Index is updated with these associations directly (455). Under Alternative B, the Auxiliary Index is updated (460)." (Column 4, line 54, - Column 5, line 2). From the context, it is clearly referencing the keyword index 405. This section does not refer to the Repository 105 which is where the complete document data is kept.

29.   Cole is not adding information from the queries to the Repository. Instead Cole adds to the keyword index. Hence, Cole does not disclose adding information from at least some of the user queries to the index of complete documents or to the enhanced web index

30.   I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

x _Einat Amitay_

Einat Amitay

Dated: 29/12/2009